Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 119 (2015) 1455 – 1462
Procedia Engineering
www.elsevier.com/locate/procedia

13th Computer Control for Water Industry Conference, CCWI 2015

# Correlation or not correlation? This is the question in modelling residential water demand pulses

Enrico Creaco[a,]*, Raziyeh Farmani[a], Lydia Vamvakeridou-Lyroudia[a], Steven G. Buchberger[b], Zoran Kapelan[a], Dragan A. Savić[a]

[a]*College of Engineering, Mathematics and Physical Sciences, University of Exeter, North Park Road, Exeter EX4 4QF - UK*
[b]*College of Engineering and Applied Science, University of Cincinnati, P.O. Box 210012 Cincinnati, OH 45221-0012 - USA*

**Abstract**

This paper presents a comparison of two different modelling approaches for the generation of residential water demand pulses as Poisson processes. Both approaches are able to preserve the mean value of daily water demand. The main difference lies in the fact that the first considers the correlation between pulse durations and intensities whereas the second neglects it. Overall, the results of the applications aimed at reproducing the measured pulses in two households show that the increase in parameterization burden associated with taking correlation into account delivers a considerable improvement in the quality of model predictions.


Peer-review under responsibility of the Scientific Committee of CCWI 2015

*Keywords:* Water demand, demand pulses, intensity, duration, correlation

**Nomenclature**

| | |
|---|---|
| $D$ | daily demand volume |
| $f$ | probability density function |
| $F$ | Weibull cumulative frequency |
| $I$ | pulse intensity |

* Corresponding author. Tel.: +44(0)1392 726421.
*E-mail address:* E.F.Creaco@exeter.ac.uk

1877-7058 
Peer-review under responsibility of the Scientific Committee of CCWI 2015
doi:10.1016/j.proeng.2015.08.1006

| | |
|---|---|
| $I^*$ | duration weighted intensity |
| $lnI$ | natural logarithm of pulse intensity |
| $lnT$ | natural logarithm of pulse duration |
| $P$ | probability in the Poisson process |
| $t$ | time |
| $T$ | pulse duration |
| $V$ | pulse volume |
| $z$ | number of generated pulses in the Poisson process |
| $\Delta t$ | time step |
| $\lambda$ | parameter of the Poisson probability distribution |
| $\mu_I$ | average value of $I$ |
| $\mu_T$ | average value of $T$ |
| $\mu_{lnI}$ | parameter of the lognormal distribution: mean of $lnI$ |
| $\mu_{lnT}$ | parameter of the lognormal distribution: mean of $lnT$ |
| $\rho$ | correlation between $T$ and $I$ |
| $\dot{\rho}$ | parameter of the lognormal distribution: correlation between $lnT$ and $lnI$ |
| $\sigma_I$ | standard deviation of $I$ |
| $\sigma_T$ | standard deviation of $T$ |
| $\sigma_{lnI}$ | parameter of the lognormal distribution: standard deviation of $lnI$ |
| $\sigma_{lnT}$ | parameter of the lognormal distribution: standard deviation of $lnT$ |
| $\mathbf{\Sigma}$ | matrix in the bivariate lognormal distribution |

## 1. Introduction

Following the "bottom-up" approach (Walski et al., 2003), the nodal demand trends used inside the water distribution system simulation models can be reconstructed by aggregating spatially and temporally the individual users' demands, which are represented as stochastic variables. Among the various residential demand generation approaches, models for the generation of water demand pulses with temporal resolution down to one second are often used. These models can be grouped in two categories with similar accuracy and performance (Blokker et al., 2009):

1 – models which use stochastic processes such as the Poisson rectangular pulse process (Buchberger and Wu, 1995; Buchberger and Wells, 1996; Guercio et al., 2001; Buchberger et al., 2003; Garcia et al., 2004; Alcocer et al., 2006; Creaco et al., 2015) or the Neyman-Scott cluster process (Alvisi et al., 2003; Alcocer-Yamanaka and Tzatchkov, 2012) to reproduce the overall water demand of the individual user without making distinction of the contributions of the various appliances of the user's;

2 – models which are able to reproduce the demand from respective micro components (i.e. by adding up the single water uses), with the aim to reconstruct the overall water demand (Blokker et al., 2010).

In the models of the first group, which are considered in the present work, the user demand is modelled by pulses whose intensity and duration are generated through prefixed probability distributions. As an example, Guercio et al. (2001) used the exponential and normal distributions for the intensity and duration respectively. Alvisi et al. (2003), Alcocer-Yamanaka et al. (2006) and Alcocer-Yamanaka and Tzatchkov (2012) used the exponential distribution for both the intensity and duration. Garcia et al. (2004), instead, used the exponential distribution only for the duration while proposing that the Weibull distribution should be used for the intensity. Buchberger et al. (2003) and Creaco et al. (2015) represent both the duration and intensity of indoor consumptions by means of the lognormal distribution. In most cases (Buchberger and Wu, 1995; Buchberger and Wells, 1996; Guercio et al., 2001; Buchberger et al., 2003; Garcia et al., 2004; Alcocer et al., 2006), pulse duration and intensity were considered to be independent random variables. However, by making calculations on the data collected on Milford households by Buchberger et al. (2003), Creaco et al. (2015) have recently shown that a non-negligible positive correlation may exist between the two variables. The Authors then stated that considering correlation helps in obtaining synthetic water demand pulses that are consistent in terms of overall daily water demand volumes, while respecting statistical properties of measured demand pulses. However, their approach still needs to be corroborated through further comparison with the approaches proposed by other authors.

This paper presents comparison of the approach proposed by Creaco et al. (2015) with the approach of Buchberger et al. (2003), which enables generation of residential demand pulses that are consistent in terms of total daily demand though considering pulse duration and intensity as independent random variables. In particular, daily demand consistence in the approach of Buchberger et al. (2003) is the result of suitable corrections made on pulse intensities.

In the following sections, first the methodology is presented, followed by applications to a literature case study.

## 2. Methodology

The two modelling approaches for the generation of demand pulse durations and intensities compared in this work are both embedded in the Poisson process for the generation of water demand pulses. In the following subsections, first the Poisson process is described, followed by the description of the two modelling approaches.

### *2.1. Poisson process*

Inside the model, time axis is scanned with a certain time resolution $\Delta t$. The probability of having $z$ generated pulses in the time interval $\Delta t$ which follows the generic time $\tau$ is described by the Poisson distribution (Buchberger and Wu, 1995):

$$P(z) = \frac{e^{-\lambda\Delta t}\left(\lambda\Delta t\right)^{z}}{z!} \quad \text{with } z = 0,1,... \tag{1}$$

where rate parameter λ represents the expected number of "events" or "arrivals" that occur per unit time.

### *2.2. First modelling approach*

In the first modelling approach, the duration $T$ and intensity $I$ of each pulse are generated using the bivariate lognormal distribution as suggested by Creaco et al. (2015). The probability density $f$ of this distribution takes on the following form:

$$f(T,I) = \frac{1}{\sqrt{(2\pi)^2|\mathbf{\Sigma}|}} \mathrm{e}^{\left\{-\frac{1}{2(1-\dot{\rho})}\left[\left(\frac{lnT-\mu_{lnT}}{\sigma_{lnT}}\right)^2+\left(\frac{lnI-\mu_{lnI}}{\sigma_{lnI}}\right)^2-\frac{2\dot{\rho}(lnT-\mu_{lnT})(lnI-\mu_{lnI})}{\sigma_{lnT}\sigma_{lnI}}\right]\right\}}, \tag{2}$$

where:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{lnT}^2 & \dot{\rho}\sigma_{lnT}\sigma_{lnI} \\ \dot{\rho}\sigma_{lnT}\sigma_{lnI} & \sigma_{lnI}^2 \end{pmatrix} \tag{3}$$

and $lnT$ and $lnI$ are the natural logarithm of $T$ and I respectively; $\mu_{lnT}$ and $\mu_{lnI}$ are the average values of $lnT$ and $lnI$ respectively; $\sigma_{lnT}$ and $\sigma_{lnI}$ are the standard deviations of $lnT$ and $lnI$ respectively; $\dot{\rho}$ is the correlation between $lnT$ and $lnI$. The adoption of this modelling approach inside the Poisson process entails considering 6 parameters: $\lambda$ in eq. (1) and $\mu_{lnT}$, $\sigma_{lnT}$, $\mu_{lnI}$, $\sigma_{lnI}$, $\dot{\rho}$ in eqs. (2) and (3). By applying the methods of moments (Hall, 2004), as suggested by Creaco et al. (2015), the 6 parameters can be related to the following parameters evaluated on the experimentally observed pulses: mean($z$), mean($T$), mean($I$), var($T$), var($I$), $\rho$, where "mean" and "var" indicate the average value and the variance respectively, and $\rho$ represents correlation between $T$ and $I$.

### 2.3. Second modelling approach

The second modelling approach, proposed by Buchberger et al. (2003), generates pulse durations $T$ and intensities $I$ as independent random variables. In this approach, first pulse durations $T$ are generated using whatever probability distribution. Subsequently, pulse intensities $I$ are generated using whatever probability distribution, calibrated on the basis of duration weighted intensity $I^*$, rather than real intensity $I$ of pulses. This artifice enables generation of consistent synthetic daily demand values even if $T$ and $I$ are uncorrelated. In particular, Buchberger et al. (2003) proposes use of the following relations for deriving mean($I^*$) and var($I^*$) from real pulses:

$$mean(I^*) = \frac{\sum T \cdot I}{\sum T} \tag{4}$$

$$var(I^*) = mean(I^{*2}) - mean(I^*)^2 = \frac{\sum T \cdot I^2}{\sum T} - \left(\frac{\sum T \cdot I}{\sum T}\right)^2 \tag{5}$$

As an alternative to eq. (5), the following equation derived from the variance of the product of two variables is here proposed:

$$var(I^*) = \frac{var(T \cdot I) - mean(I)^2 var(T)}{mean(T)^2 + var(T)} \tag{6}$$

Two options then exist for the application of this modelling approach: option 1 based on eqs. (4) and (5) and option 2 based on eqs. (4) and (6).

Though this modelling approach can be applied to a number of probability distributions, it will be applied hereinafter using the mono-variate lognormal distribution for the generation of both $T$ and $I$. This was done in order to facilitate comparison with the first modelling approach, in which the lognormal distribution is the marginal probability distribution for either $T$ or $I$. Adoption of this modelling approach inside the Poisson process then entails considering 6 parameters: $\lambda$ in eq. (1) relative to pulse arrivals and $\mu_{lnT}$, $\sigma_{lnT}$, $\mu_{lnI}$, $\sigma_{lnI}$ relative to $T$ and $I$ generated through mono-variate lognormal distributions. By applying the methods of moments (Hall, 2004), the 6 parameters can be related to the following parameters evaluated on the experimentally observed pulses: mean($z$), mean($T$), mean($I^*$), var($T$), var($I^*$).

## 3. Applications

### 3.1. Case study

By making calculations on data collected during an experimental campaign in some households in Milford, Buchberger et al. (2003) were able to reconstruct, with one-second time step resolution, the water demand pulses that were taking place in these households in the period from April to October 1997. The data made available by the Authors concern pulse duration $T$, intensity $I$ and volume $V=T \cdot I$.

As case study in this work, the indoor water demand pulses recorded in two of the households, namely households 1 and 2. For these households, the months of June and April were selected respectively. The basic statistical parameters of measured water consumption variables $z$, $T$, $I$, $V$, $D$ (= daily demand value) and $\rho$ in the two tests are reported in Tables 1 and 2 respectively.

The modelling framework of this paper was aimed at comparing the two different modelling approaches described in section 2. In particular, model A was developed following the first modelling approach (section 2.2). Models B and C were developed following the second modelling framework, considering options 1 and 2 described in section 2.3 respectively.

Overall, applications consisted in 3 phases for each of the two tests and for each of the three models:

phase 1 – parameter assessment;

phase 2 – generation of synthetic water demand pulses;
phase 3 – analysis of the results of the models and comparison with the observed data.

### 3.2. Results

Phase 1

The results of phase 1 for models A, B and C are reported in Table 3, 4 and 5. In particular, Table 3 report the calibrated values of $\lambda$, parameter relative to pulse arrival times, in the 12 time slots for the two tests. The values of $\lambda$ are common to the three models, which only differ in the way $T$ and $I$ are assessed. Tables 4 and 5 report parameters $\mu_{lnT}$, $\sigma_{lnT}$, $\mu_{lnI}$, $\sigma_{lnI}$ and $\dot{\rho}$ for tests 1 and 2 respectively.

Table 1 – Test 1. Basic statistical parameters of water consumption variables $z$, $T$, $I$, $V$, $D$ and $\rho$ derived from the measured pulses and from the pulses generated by the models.

| | mean ($z$) [$s^{-1}$] | mean ($T$) [sec] | var ($T$) [$sec^2$] | mean ($I$) [L/s] | var ($I$) [$L^2/s^2$] | mean ($V$) [L] | var ($V$) [$L^2$] | mean ($D$) [L/day] | cov ($I$,$T$) [$L^2$] | $\rho$ [-] |
|---|---|---|---|---|---|---|---|---|---|---|
| measured | 0.00099 | 36 | 4624 | 0.086 | 0.0029 | 4.86 | 137 | 417 | 1.79 | 0.49 |
| model A | 0.00099 | 36 | 4571 | 0.086 | 0.0030 | 4.92 | 365 | 422 | 1.84 | 0.50 |
| model B | 0.00099 | 36 | 5015 | 0.137 | 0.0050 | 4.91 | 127 | 421 | -0.01 | 0.00 |
| model C | 0.00099 | 36 | 5005 | 0.136 | 0.0085 | 4.87 | 147 | 417 | -0.01 | 0.00 |

Table 2 – Test 2. Basic statistical parameters of water consumption variables $z$, $T$, $I$, $V$, $D$ and $\rho$ derived from the measured pulses and from the pulses generated by the models.

| | mean ($z$) [$s^{-1}$] | mean ($T$) [sec] | var ($T$) [$sec^2$] | mean ($I$) [L/s] | var ($I$) [$L^2/s^2$] | mean ($V$) [L] | var ($V$) [$L^2$] | mean ($D$) [L/day] | cov ($I$,$T$) [$L^2$] | $\rho$ [-] |
|---|---|---|---|---|---|---|---|---|---|---|
| measured | 0.00054 | 56 | 12743 | 0.106 | 0.0074 | 9.45 | 505 | 442 | 3.50 | 0.36 |
| model A | 0.00054 | 56 | 11986 | 0.106 | 0.0074 | 9.46 | 1369 | 441 | 3.51 | 0.37 |
| model B | 0.00054 | 56 | 12462 | 0.168 | 0.0109 | 9.48 | 577 | 442 | 0.04 | 0.00 |
| model C | 0.00054 | 56 | 12462 | 0.168 | 0.0091 | 9.47 | 544 | 442 | 0.04 | 0.00 |

Table 3 – Values of $\lambda$ for all models and for both tests.

| test | 0-2 | 2-4 | 4-6 | 6-8 | 8-10 | 10-12 | 12-14 | 14-16 | 16-18 | 18-20 | 20-22 | 22-24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7.87e-5 | 7.87e-5 | 1.94e-4 | 2.91e-3 | 1.08e-3 | 1.85e-3 | 8.33e-4 | 8.43e-4 | 1.45e-3 | 1.26e-3 | 9.31e-4 | 4.21e-4 |
| 2 | 6.48e-5 | 2.78e-5 | 1.06e-3 | 9.91e-4 | 5.79e-4 | 3.56e-4 | 1.30e-4 | 3.33e-4 | 1.02e-3 | 1.03e-3 | 7.45e-4 | 1.57e-4 |

Table 4 – Test 1 - Values of parameters $\mu_{lnT}$, $\sigma_{lnT}$, $\mu_{lnI}$, $\sigma_{lnI}$ and $\dot{\rho}$ for the models.

| model | $\mu_{lnT}$ | $\sigma_{lnT}$ | $\mu_{lnI}$ | $\sigma_{lnI}$ | $\dot{\rho}$ |
|---|---|---|---|---|---|
| A | 2.81 | 1.24 | -2.62 | 0.58 | 0.65 |
| B | 2.81 | 1.24 | -2.11 | 0.49 | - |
| C | 2.81 | 1.24 | -2.18 | 0.62 | - |

Table 5 – Test 2 - Values of parameters $\mu_{lnT}$, $\sigma_{lnT}$, $\mu_{lnI}$, $\sigma_{lnI}$ and $\dot{\rho}$ for the models.

| model | $\mu_T$ | $\sigma_T$ | $\mu_I$ | $\sigma_I$ | $\dot{\rho}$ |
|---|---|---|---|---|---|
| A | 3.22 | 1.27 | -2.50 | 0.71 | 0.52 |
| B | 3.22 | 1.27 | -1.94 | 0.57 | - |
| C | 3.22 | 1.27 | -1.92 | 0.53 | - |

Phase 2

The models calibrated in phase 1 were then applied in order to create synthetic demand pulses for one month for each test. In order to account for the influence of the random seed, each one month long pulse generation was repeated 100 times.

As an example, Fig. 1 shows the single realization of the simulated total demand for a typical day at the scale of 1 sec, obtained by using model C. As expected, the figure shows a higher concentration of pulses in the morning and in the late afternoon, when the household occupants usually get up and get back home after work respectively. Very few pulses are instead generated at nighttime.

Phase 3

A first analysis was made concerning the basic statistical parameters of water consumption variables $z$, $T$, $I$, $V$ and $\rho$ derived from the pulses generated by models A-C, in comparison with those of the measured pulses in both tests (see Tables 1 and 2). These tables show that all the models reproduce well mean($z$), mean($T$), var($T$), mean($V$) and mean($D$). Compared to models B and C, model A better preserves mean($I$), var($I$), cov($T$,$I$) and $\rho$. The better consistency of model A in terms of mean($I$) and var($I$) has to be ascribed to the fact that pulses in model A are calibrated with reference to real intensities. In models B and C, pulse durations are, instead, calibrated with reference to duration weighted intensities. The better consistency of model A in terms of cov($T$,$I$) and $\rho$ is due to the fact that model A is the only one that considers correlation between $T$ and $I$. Models B and C are more consistent in terms of var($V$), as a result of use of eqs. (5) and (6) respectively. In fact, these equations have the effect of getting var($V$) preserved.

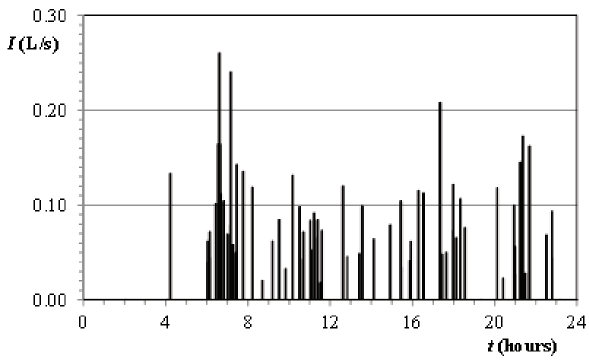


Fig. 1. Model A - single realization of the simulated total demand for a typical day at the scale of 1 sec.

Another analysis was then carried out to examine the consistency of the synthetic water demand pulses generated by means of the models with the measured water demand pulses in terms of overall daily water demand volume $D$. In particular, for each test the total synthetic water demand volume $D$ was calculated for each day in the generic 1 month long pulse generation. Then, the cumulative frequency curve was constructed reporting, for each value of $D$, the Weibull cumulative frequency $F$ of days in the month that feature an overall daily water demand volume lower than or equal to $D$. Since each model application comprises 100 one-month long pulse generations, a band of synthetic cumulative frequencies was then obtained for each test. For each test, the band upper envelope (BUE), lower envelope (BLE) and mean value (BMV) of the 100 cumulative frequency curves were determined for all the models. The cumulative frequency of the measured daily water demand volume (ECF) was also calculated.

The graphs in Figure 2 report, for each test, BUE, BLE and BMV obtained using the various models as well as ECF. Analysis of the graphs shows that, for test 1, the BMV obtained with model A (model that takes account of the mutual dependence of pulse intensity and duration by means of the bivariate distribution) follows ECF much more closely than those obtained with models B and C (model that neglects the mutual dependence of pulse intensity and duration). This behavior is noticed over all the values of daily demand *D*. Furthermore, with regard to test 1, all the data points of ECF but one lie inside the band of cumulative frequency obtained with model A. A much higher number of data points of ECF lie, instead, outside the bands obtained by means of models B and C.

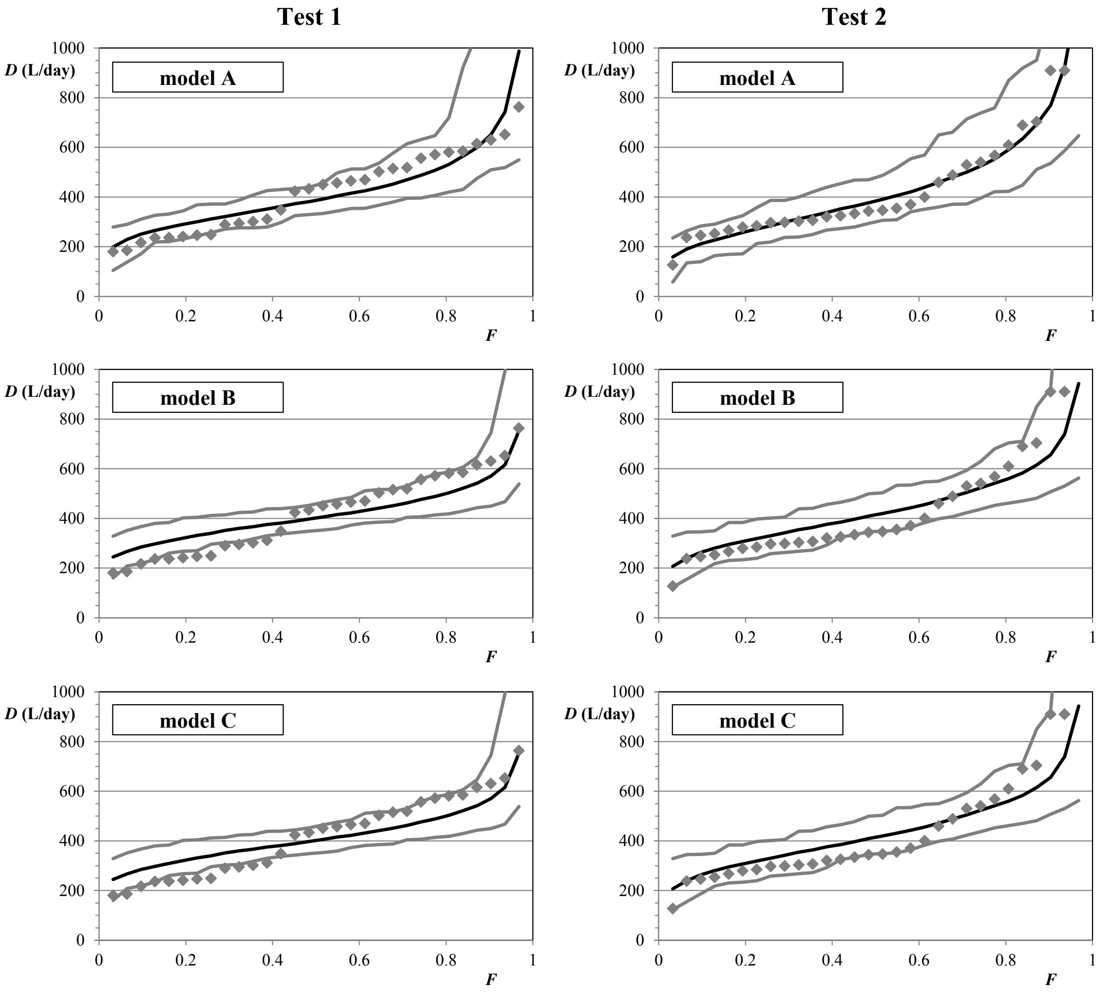


Fig. 2. Upper and lower envelopes (grey lines) and mean value (black line) of the band of Weibull cumulative frequencies *F* of daily water demand *D* produced by models A, B and C in comparison with the daily water demand cumulative frequency calculated starting from the measured data (dots).

Similar remarks can also be made for test 2, though in this case the worse performance of models B and C are less evident. This may be ascribed to the fact that the correlation in real pulses is significantly larger in test 1

(=0.49 see Table 1) than in test 2 (=0.36). The benefits derived from taking correlation into account are therefore more evident in test 1 than test 2.

Overall, results indicate that model A performs better than models B and C. This corroborates the findings of Creaco et al. (2015), in that taking account of correlation enables better preservation of pulse statistical properties and better representation of total daily demands. No significant differences were remarked between models B and C. This means that the options described in section 2.3 for obtaining the average value and the standard deviation of the duration weighted intensity in the second modelling framework lead to similar results.

## 4. Conclusions

Two modelling approaches were compared, which can be embedded in the Poisson process for generating residential demand pulses. The first approach considers correlation between pulse duration and intensity whereas the second treats the two variables as independent random variables. Results of applications showed that considering correlation enables better preservation of the statistical properties of the pulses, which are also more consistent in terms of total daily demand. The benefits derived from considering correlation in the modelling are more evident in case studies where the duration and intensity of the real pulses are strongly correlated.

A drawback of the models with correlation lies in an increase in the parameterization burden, due to the presence of one extra parameter (i.e., the duration/intensity correlation) that needs to be calibrated. The absence in the scientific literature of well established procedures that can be used for parameter assessment when data concerning real demand pulses are not available must also be highlighted.

## Acknowledgements

This study was carried out as part of the ongoing “iWIDGET” project (Grant Agreement No 318272), which is funded by the European Commission within the 7th Framework Programme.

## References

Alcocer-Yamanaka, V.H., Tzatchkov, V., and Buchberger, S.G., 2006. Instantaneous Water Demand Parameter Estimation from Coarse Meter Readings. In: Proceedings of the Water Distribution Systems Analysis Symposium 2006, pp. 1-14.

Alcocer-Yamanaka, V.H., and Tzatchkov, V.G., 2012. Modeling of Drinking Water Distribution Networks Using Stochastic Demand. Water Resources Management, 26, 1779-1792.

Alvisi, S., Franchini, M., and Marinelli, A., 2003. A stochastic model for representing drinking water demand at residential level. Water Resources Management, 17 (3), 197–222.

Blokker, E.J.M., Buchberger, S.G., Vreeburg, J.H.G., and van Dijk, J.C., 2009. Comparison of Water Demand Models: PRP and SIMDEUM Applied to Milford, Ohio, Data. In: Proceedings of the Water Distribution Systems Analysis 2008, pp. 1-14.

Blokker, E.J.M., Vreeburg, J.H.G., and van Dijk, J.C., 2010. Simulating residential water demand with a stochastic enduse model. Journal of Water Resources Planning and Management, 136 (1), 375–382.

Buchberger, S.G., Carter, J.T., Lee, Y.H., and Schade, T.G., 2003. Random Demands, Travel Times and Water Quality in Dead-ends, prepared for American Water Works Association Research Foundation, AwwaRF Report No. 294, 2003, 470 pages.

Buchberger, S.G., and Wells, G.J., 1996. Intensity, duration and frequency of residential water demands. Journal of Water Resources Planning and Management, 122 (1), 11–19.

Buchberger, S.G., and Wu, L., 1995. Model for instantaneous residential water demands. Journal of Hydraulic Engineering, 121 (3), 232–246.

Creaco, E. , Farmani, R., and Kapelan, Z., Vamvakeridou-Lyroudia, L., Savic, D., 2015. Considering the mutual dependence of the pulse duration and intensity in models for generating residential water demand. Under review on Journal of Water Resources Planning and Management.

Garcıa, V.J., Garcıa-Bartual, R., Cabrera, E., Arregui, F., and Garcıa-Serra, J. 2004. Stochastic model to evaluate residential water demands. Journal of Water Resources Planning and Management, 130 (5), 386–394.

Guercio, R., Magini, R., and Pallavicini, I., 2001. Instantaneous residential water demand as stochastic point process. In: Water resources management, C. A. Brebbia et al. eds., WIT Press, Southampton, U.K., 129-138.

Hall, A.R., 2004. Generalized Method of Moments. Advanced Texts in Econometrics, OUP Oxford, ISBN: 0198775202, 9780198775201.

Johnson, R.A. and Bhattacharyya, G.K., 1992. Statistics: Principles and Methods. 2nd Edition. Wiley, 1992.

Walski, M., Chase, D., Savic, D., Grayman, W., Beckwith, S. and Koelle, E. 2003. Advanced water distribution modelling and management. Waterbury, CT: Haestad Press.